



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

What Matters for Agricultural Trade? Assessing the Role of Trade Deal Provisions using Machine Learning

Stepan Gordeev, Jeremy Jelliffe, Dongin Kim, and Sandro Steinbach

Selected presentation for the International Agricultural Trade Research Consortium's (IATRC's) 2023 Annual Meeting: The Future of (Ag-) Trade and Trade Governance in Times of Economic Sanctions and Declining Multilateralism, December 10-12, 2023, Clearwater Beach, FL.

Copyright 2023 by Stepan Gordeev, Jeremy Jelliffe, Dongin Kim, and Sandro Steinbach. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

What Matters for Agricultural Trade?

Assessing the Role of Trade Deal Provisions using Machine Learning

Stepan Gordeev

Agricultural and Resource Economics
University of Connecticut

Jeremy Jelliffe*

Economic Research Service
United States Department of Agriculture

Dongin Kim

Agribusiness and Applied Economics
North Dakota State University

Sandro Steinbach

Agribusiness and Applied Economics
North Dakota State University

IATRC Annual Meetings
December 10-12, 2023

Disclaimer: The findings and conclusions in this abstract submission are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.



1. Preferential Trade Agreements (PTAs) increasingly dominate international trade governance
 - Multilateral negotiation through WTO -> Regional negotiation through PTA
 - Only 50 PTAs in 1990s -> 354 active PTAs in 2023
 2. PTAs are becoming increasingly deep and varied in contents
 - PTAs seek to improve market access
 - Tariff is already low -> number of **non-tariff provisions** rises
- This paper:
- 1) What **PTA provisions** affect agricultural trade?
 - 2) What **factors** determine containing such provisions in PTAs?
- ✓ Exploit **machine learning** to answer these questions.



Background: Number of PTAs Has Skyrocketed

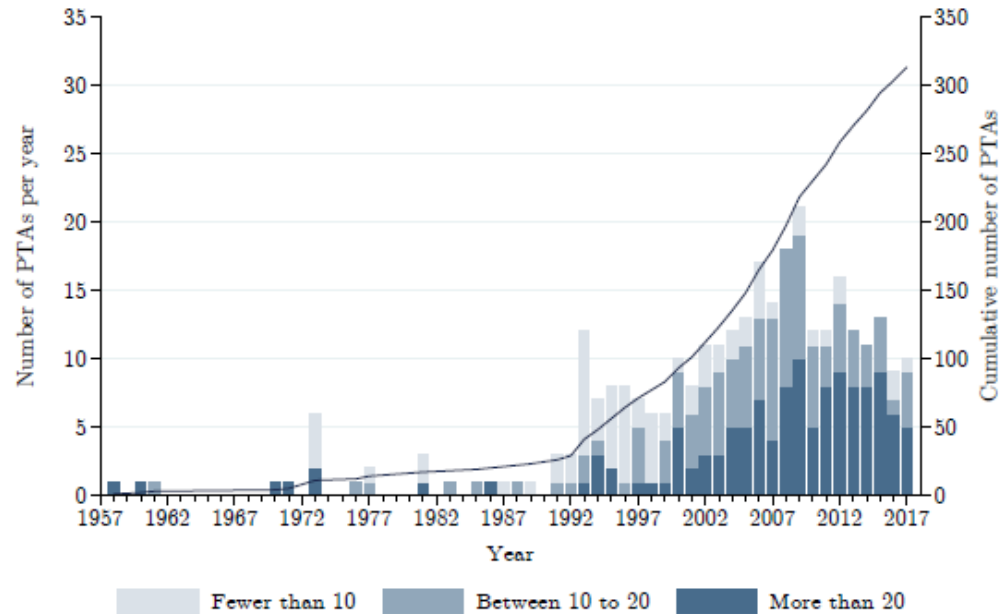


Figure 1. The Increase of Active PTAs and Policy Areas

Source. Author's calculation based on Hofmann et al (2017).

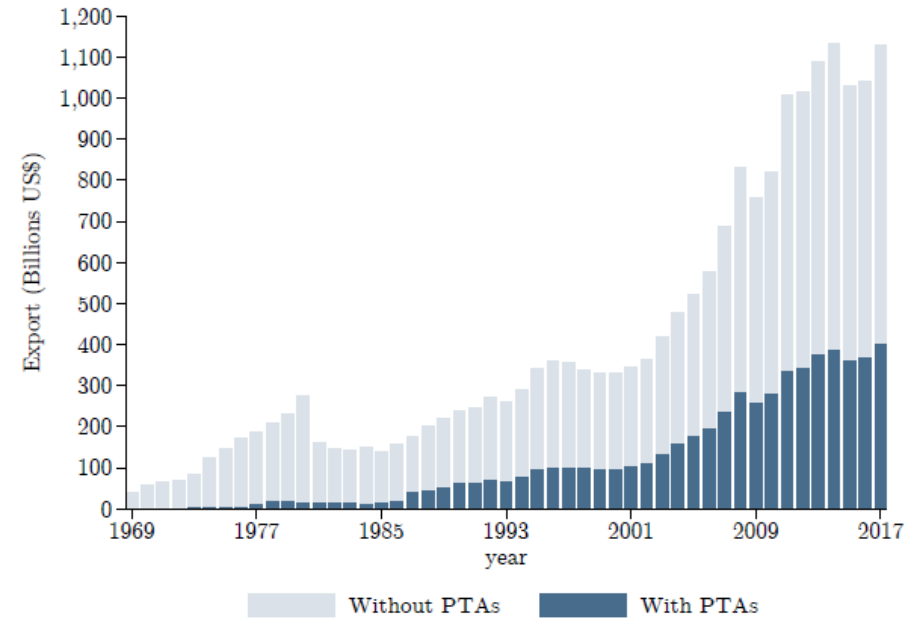


Figure 2. Agricultural Export with PTA Partners

Source. Author's calculation based on Hofmann et al (2017) and UN Comtrade.



Background: PTA Contents are Becoming More Complex and Diverse

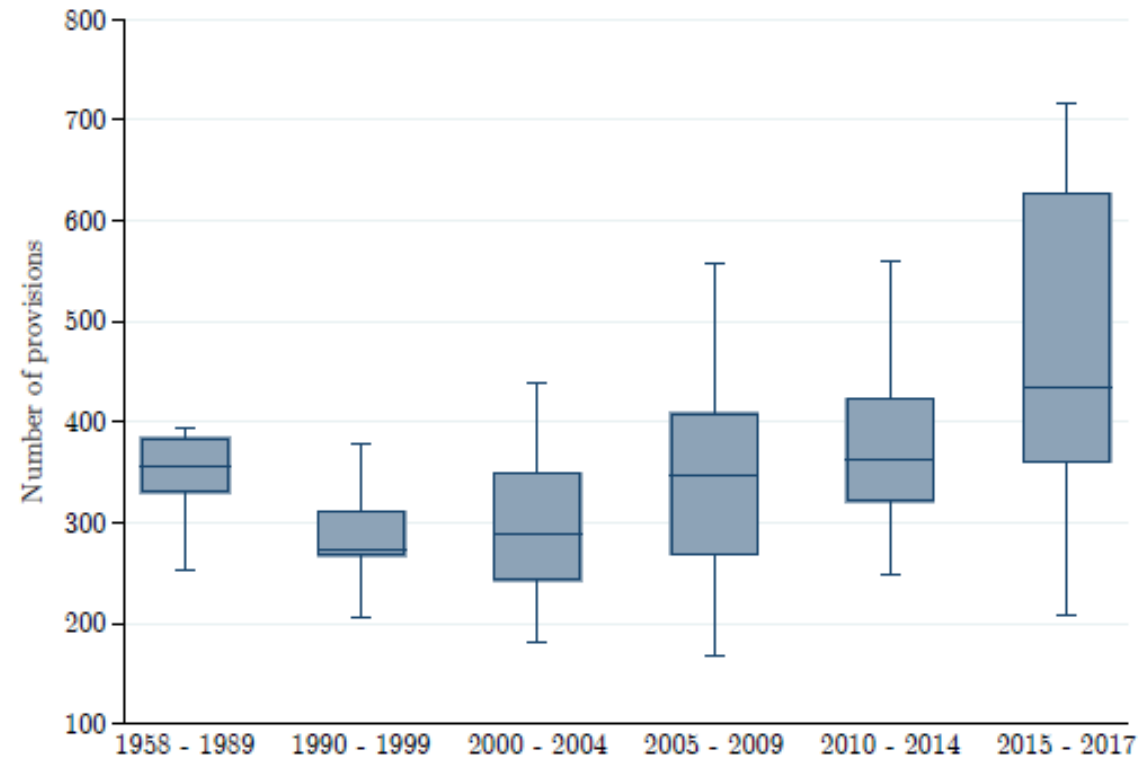


Figure 3. Number of Provisions of Enforced PTAs

Source. Author's calculation based on Mattoo et al. (2020).



1. Effective PTA Provisions

- Explore a specific provision area: Raess et al. (2018), Kucik (2012), Lechner (2016)
- PTA depth: Gamsö and Grosse (2021), Mattoo et al. (2020)
- Lasso to identify PTA provisions: Breinlich et al. (2022)

2. Determinants of PTA design and formation

- Explore a specific mechanism: Baldwin and Jaimovich (2012), Mansfield and Milner (2012), Baccini and Urpelainen (2014)
- Compare several determinants: Baier and Bergstrand (2004), Bergstrand et al. (2016)

- This paper: Combines the idea of **identifying the binding PTA provisions** on agricultural trade and investigating the **determinants** of the effective PTAs to understand 'what matters for agricultural trade.'



- 1. Identifying** the effective PTA provisions on agricultural trade
 - Utilize **Plug-in Lasso** regularized regression
 - Address the challenge of multicollinearity arising from numerous correlated policy variables
- 2. Determining** the factors influencing the identified provisions
 - Apply **Random Forests** method
 - Excel in capturing non-linearity and interactive effects among potential factors



Gravity Model

✓ Three-way Gravity Model:

$$X_{ijt} = \exp(\alpha_{it} + \gamma_{jt} + \delta_{ij} + PTA'_{ijt}\beta') * \mu_{ijt}$$

- X_{ijt} is bilateral agricultural trade from country i to j at year t
 - α_{it} and γ_{jt} encompass the time-varying country dummy variables
 - δ_{ij} denotes the set of country-pair fixed effects
 - PTA'_{ijt} represents the vector of PTA provisions
- ✓ Poisson Pseudo Maximum Likelihood (**PPML**) to estimate β' (Silva and Tenreyro, 2006)
- Challenge: Identify the trade effects of (many) correlated provisions
 - Solution: Reduce the dimension $PTA'_l \subseteq PTA'$



Plug-In Lasso Regularized Regression Primer

- ✓ Use **Lasso regression** to identify major PTA provisions
 - Penalizing extra variables that do not significantly improve the model specification.

$$\arg \min_{\beta_l} \underbrace{RSS}_{\text{Penalty Term}} + \lambda \underbrace{\sum_{l=1}^m |\beta_l|}_{\text{Penalty Term}}$$

Additional variable(s): ↓ ↑

- ✓ **PPML version of Plug-in Lasso** suggested by *Breinlich et al. (2022)*
 - Apply *Belloni et al. (2016)*'s Plug-in method to account for heteroskedasticity of each variable
 - Offer the most restrictive selection results among other techniques
- ✓ Why Lasso?
 - Need only $m+1$ to train the data -> Avoid overfitting + multicollinearity issue
 - Can set the specific regression form -> Gravity theory-consistent



Random Forests Basics

✓ Procedure

- Grow large number of decision trees, each using a bootstrap sample
- Within each decision tree:
 - Each node splits the data on one of the determinant variables (D_i), optimizing some measure of fit
 - Bottom “leaves” classify the observation into groups 0 or 1
- Classify an observation: All trees classify it, majority wins

✓ Why Random Forests?

- Performs well when number of variable (D_i) is high
- Adapts to non-linearities and interactions in the data
- Well-developed variable importance measures

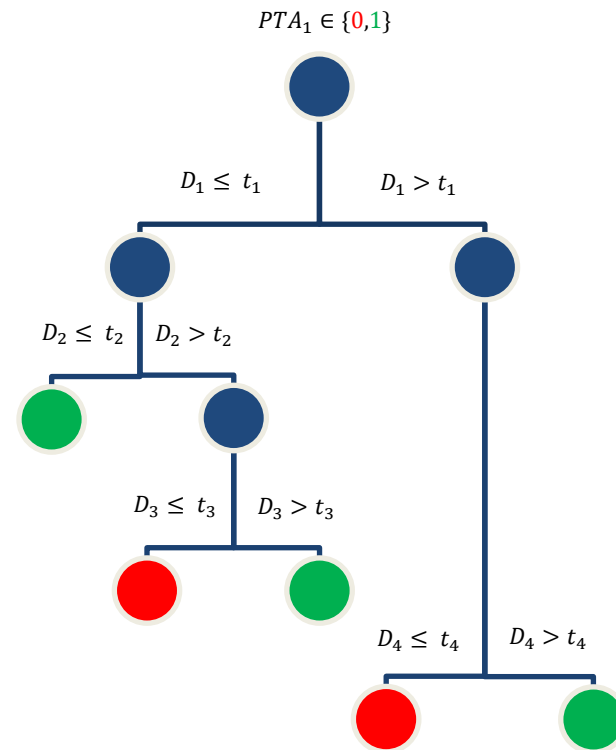


Figure 4. Decision Tree Example



- ✓ Agricultural trade flows
 - Obtain product-level export data from **UN Comtrade**
 - Aggregated all products under SITC codes 00-09.
 - Covers 213 exporters and 260 importers from 1968 to 2017

- ✓ PTA Provisions
 - **Deep Trade Agreements** from World Bank
 - 937 provisions for 282 active PTAs
 - Used 298 indicator variables of the essential provision classified by Mattoo et al. (2022)
 - Excluded observations for countries associated with the 35 unmapped-PTAs
 - Used DESTA database to drop the observation with the records of expired PTAs



✓ Determinants

- Identify 291 economic, geographic, and political factors based on PTA formation literature (Baldwin and Jaimovich, 2012; Mansfield and Milner, 2012; Baccini, 2019)
 - Factors found to be important determinants of overall PTA formation (e.g., macroeconomic variables, PTA contagion, domestic politics , etc.)
 - Factors related to particular PTA provision policy areas (e.g., innovation, energy use, labor markets, sectoral variables, etc.)
- Collect and compile data
 - **Economy, society:** Penn World Tables, World Development Indicators
 - **Proximity, culture:** CEPII Gravity, CEPII Language, GeoDist, UNCTAD
 - **Trade, FDI:** UN Comtrade, WITS, IMF CDIS
 - **Politics:** Database of Political Institutions, Worldwide Governance Indicators
- Construct final dataset at **country-pair** level



The Binding PTA Provisions

| | PPML (1) | Lasso (2) | PPML (3) | PPML (4) |
|---|---------------------|--------------|----------------------|----------------------|
| PTA | 0.307*** (0.071) | | | -0.005 (0.064) |
| CP.23 – Contains provisions that promote transparency | | 0.222 | 0.236*** (0.080) | 0.236*** (0.080) |
| ET.15 – Requires phase out of existing export taxes | | 0.03 | 0.047 (0.083) | 0.047 (0.083) |
| ET.18 – Prohibits an increase in the rate of existing export taxes | | 0.109 | 0.312*** (0.071) | 0.314*** (0.072) |
| IPR.58 – Designates that any parties meeting a particular specification may use a GI without registering | | -0.012 | -0.376*** (0.095) | -0.376*** (0.142) |
| IPR.88 – Industrial Design: Provides minimum term of protection | | -0.060 | -0.327*** (0.095) | -0.327*** (0.095) |
| MoC.37 – Excludes ‘good faith and non-discriminatory application of its laws’ governing capital account regulations | | 0.099 | 0.250* (0.140) | 0.249* (0.140) |
| MoC.38 – Contains country annexes with specific transfer reservations by individual parties | | 0.021 | 0.200*** (0.067) | 0.201*** (0.067) |
| STE.43 – Includes any other specific discipline for certain sectors or objectives | | 0.630 | 0.881*** (0.100) | 0.883*** (0.102) |
| TBT.4 – Mutual recognition in force on integrating standards | | 0.025 | 0.621*** (0.178) | 0.621*** (0.178) |
| Observation | 368,227 | 368,227 | 368,227 | 368,227 |
| Pseudo R^2 | 0.939 | | 0.941 | 0.941 |

Note: Heteroskedasticity-robust standard errors are in the parentheses, clustered by exporter-importer pair. For the estimated parameters, the semi-elasticity is given by $100 * (\exp(\beta) - 1) \% \text{ expect } \text{Log}(1 + \text{Tariff})$.



Identifying the Determinants: RF Prediction Performance

- ✓ Out-of-Bag (OOB)
 - Instances that were not included in the training set for a particular tree.
- ✓ OOB Misclassification Rate
 - Dividing the total number of misclassified instances by the total number of OOB instances.
- Maximum OOB rate: 14% (*excellent fit of the data)

| Provision | OOB Misclassification |
|-----------|-----------------------|
| CP.23 | 0.14 |
| ET.15 | 0.11 |
| ET.18 | 0.14 |
| IPR.58 | 0.02 |
| IPR.88 | 0.07 |
| MoC.37 | 0.04 |
| MoC.38 | 0.04 |
| STE.43 | 0.07 |
| TBT.4 | 0.03 |

Note: The OOB misclassification error is calculated by the fraction of country-pair-year observations with PTAs for which the estimated RF predicts the presence of the particular PTA provision incorrectly, computed out-of-bag.



Major Determinants for the Binding PTA Provisions

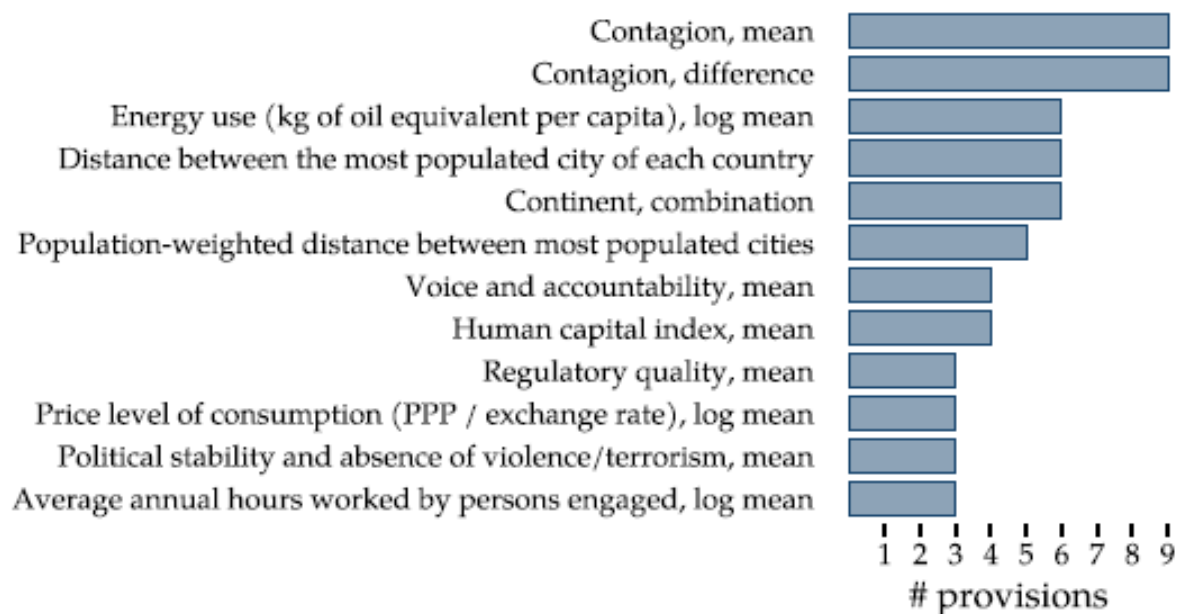


Figure 5. Important Determinants of Binding PTA Provisions

Note. The figure lists the major determinants of binding PTA provisions. “# provisions” denotes the number of binding provisions for which each determinant is significantly important (the p-value of the permutation variable importance measure is below 0.05). **Only the determinants relevant to three or more provisions are displayed.**

✓ Contagion

- The existing trade agreements between countries, including the associated third parties, influence the decision to sign a PTA and determine the content of the new trade deal.

✓ Others

- Similarities in income level, political/legal maturity, and the size of the economy are important factors.



Interpreting the Determinants

| | <u>Hurdle Model</u> | | |
|--|----------------------|-----------------------|---------------------|
| | Count | Any | Count (≥ 1) |
| Contagion, mean | 0.057*** (0.009) | 36.669*** (10.116) | 0.589*** (0.205) |
| Contagion, difference | -0.009 (0.014) | -23.057*** (6.384) | -0.123 (0.168) |
| Energy use (kg of oil equivalent per capita), log mean | 0.195*** (0.019) | 0.525*** (0.080) | 0.138*** (0.027) |
| Distance between the most populated city | 0.774*** (0.338) | -3.518** (1.534) | 0.000*** (0.000) |
| Population-weighted distance between most-populated cities | -0.941*** (0.344) | 3.555** (1.550) | 0.000*** (0.000) |

Note: Column “Count” presents the results of a Poisson regression on the number of agriculture-relevant provisions in a PTA on the determinants identified by the RFs to be important (p -value < 0.05) for over half of the selected provisions. Column “Any” presents the results of a logistic regression on whether any agriculture-relevant provisions are present. Column “Count (≥ 1)” presents results of a Poisson regression on the count of agriculture-relevant provisions in a PTA, conditional on there being some.



Conclusion

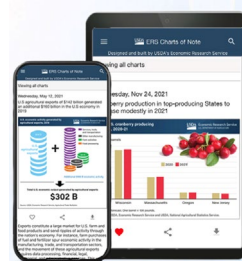
- ✓ Findings from Lasso and RF models provide new insights into PTA provisions driving agricultural trade and factors influencing their inclusion in trade deals.
- ✓ The **Lasso model** identifies the most impactful PTA provisions for agricultural trade, including competition policy, export taxes, IPR, movement of capital, state enterprises, and TBT.
- ✓ The **RF model** determines the most influential factors for including these provisions, highlighting contagion (mean) and energy use as positively associated and statistically significant, while contagion (difference) shows an inverse association.
- ✓ In summary, agricultural trade provisions are more likely to be shared between similar or aligned countries, while misalignment leads to divergence in these provisions.



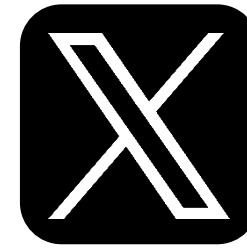
Like, Share, & Follow ERS



www.ers.usda.gov



[www.ers.usda.gov/
data-products/charts-of-note](http://www.ers.usda.gov/data-products/charts-of-note)



@USDA_ERS



[linkedin.com/company/
usda-economic-research-service](https://linkedin.com/company/usda-economic-research-service)

Subscribe to Weekly E-mail Notifications: www.ers.usda.gov/subscribe

Learn About Careers at ERS: www.ers.usda.gov/about-ers/careers-at-ers





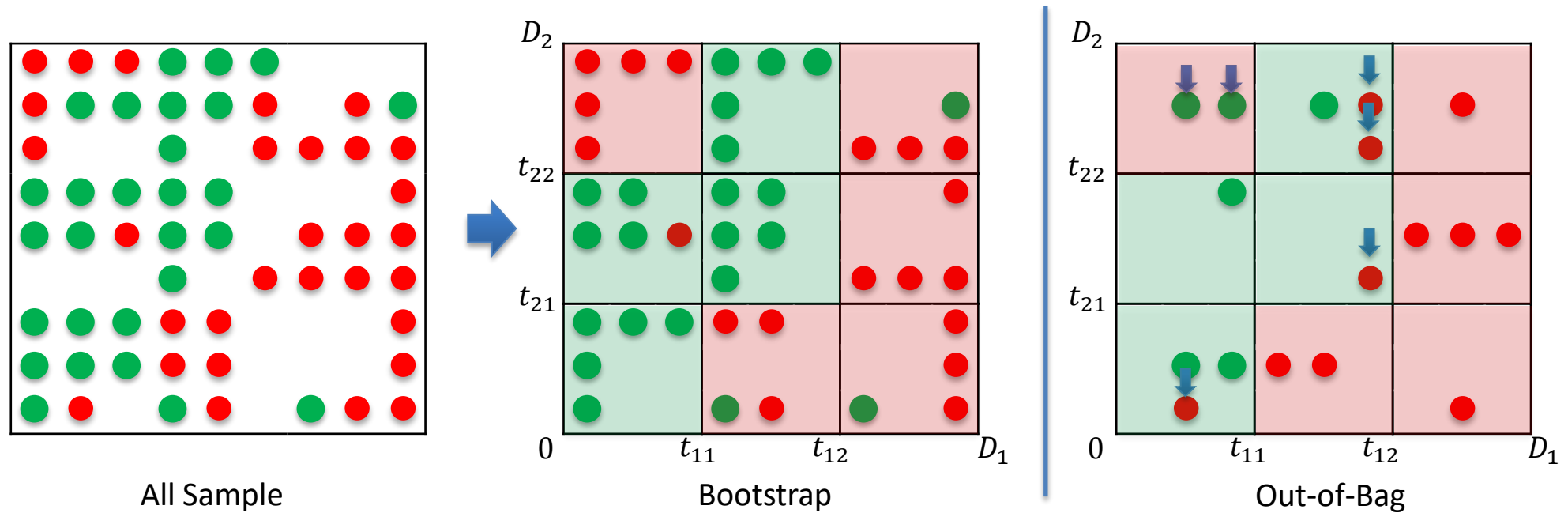


Figure 6. Example of Bootstrapping and Out-of-Bag Errors



Supp: P-Values of the Major Determinants of Binding Provisions

| Determinant | CP 23 | ET 15 | ET 18 | IPR 58 | IPR 88 | MoC 37 | MoC 38 | SE 43 | TBT 4 |
|---|----------|----------|----------|-----------|-----------|-----------|-----------|----------|----------|
| Contagion, mean | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Contagion, difference | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Energy use (kg of oil equivalent per capita), log mean | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.20 | 0.00 | 0.00 |
| Distance between the most populated city of each country | 0.00 | 0.00 | 0.00 | 0.82 | 0.00 | 0.04 | 0.64 | 1.00 | 0.03 |
| Continent, combination | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.19 | 1.00 | 0.00 |
| Population-weighted distance between most populated cities | 0.00 | 0.00 | 0.00 | 0.94 | 0.00 | 0.07 | 0.74 | 1.00 | 0.00 |
| Voice and accountability, mean | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.17 | 0.10 | 0.00 | 0.23 |
| Human capital index, mean | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.87 | 0.28 | 0.00 | 1.00 |
| Regulatory quality, mean | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.98 | 0.40 | 1.00 | 0.93 |
| Price level of consumption (PPP / exchange rate), log mean | 0.80 | 0.00 | 0.00 | 0.99 | 0.67 | 1.00 | 0.24 | 0.00 | 1.00 |
| Political stability and absence of violence/terrorism, mean | 0.01 | 0.01 | 0.00 | 1.00 | 1.00 | 1.00 | 0.23 | 1.00 | 1.00 |
| Average annual hours worked by persons engaged, log mean | 0.02 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 0.66 | 1.00 | 1.00 |

Note. The table shows p-values of variable importance for the major determinants of binding PTA provisions fostering agricultural trade. The order of determinants follows Figure 2, and only the determinants relevant for three or more provisions are displayed. Individual provision columns use abbreviated versions of provision names in Table 3 and display the p-values of the permutation variable importance measures for selected determinants. The p-value color coding indicates **p-value < 0.01**, **p-value < 0.05**, and **p-value < 0.1**.

